Individual Assignment 1: Visualizing Text

Chris Anderson
PREDICT 455 Section 56

Summary and Problem Definition

A political analysis company aims to understand and compare the topics discussed by each candidate at the first 2016 presidential debate. Reactionary analysis can potentially be swayed by a candidate's temperament, emphatic moments throughout the debate, or portrayal in the media. While the telecast of the debate was a whirlwind of misdirection and interruptions, using data visualization tools to analyze the transcript can provide an objective presentation of each candidate's primary talking points.

Methods

A transcript of the debate was obtained from the website of The American Presidency Project (Peters & Woolley, 2016). The original format of the transcript was not raw text, so some manipulation was required to convert the text to a format suitable for analysis and visualization. Microsoft Excel formulas and filtering were used to group the text by speaker and ultimately create two text files. The first contained every word spoken by Democratic candidate Hilary Clinton, and the second contained the same for Republican candidate Donald Trump. Then, R was used to cleanse the text of punctuation marks and spaces, identify whole words, and create a table showing the frequency for each word. The frequency table was extracted into a .csv file to be used in Tableau.

Programming Overview

In R, the *readr*, *tm*, and *stringr* packages were used to analyze the raw text files. First, the Hilary Clinton file was brought into R and all text was converted to lowercase letters for consistency. Since the source was a transcription of spoken words, it was filled with contractions, which cause issues when parsing the text into individual words. To combat this issue, common contractions were replaced with the words they represent. Then, functions were used to split each line into individual words and count the number of times each word appeared in the text. Finally, a table was constructed with one row for each unique word along with its frequency. The same process was repeated for the Donald Trump text file. The

frequency tables were subsequently extracted from R and used as data sources in Tableau. Additional data sources used in Tableau include a list of common words to exclude from the analysis and a list of key topics from the debate. Using links between these sources, a word cloud was built for each candidate, as well as a horizontal bar chart to comparing their usage of selected words. The word cloud graphics were not programmed to include a specific number of words, but rather to contain all words mentioned at least 10 times.

Results

The resulting graphic presents an interesting comparison between each candidate's major themes during the debate. In a general sense, the sheer number of words included in Donald Trump's word cloud suggests that his message was significantly more far-reaching, while Hilary Clinton limited her scope to fewer key topics. Both candidates referred to their opposition by name repeatedly, indicative of the combative nature of the debate. There are some similarities, though, in the use of the common political phrases like "people", "country", "tax", and "jobs". The bar chart presents a direct comparison of the frequency of certain words. The words "business", "economy", "justice", "plans", and "together" were spoken far more frequently by Hilary Clinton, while the words "Chicago", "defend", "Mexico", "money", "politicians", and "war" were spoken more often by Donald Trump. These graphics illustrate the important points of each candidate's message. Management at the political analysis company should use these results to present an objective review of the content discussed during the debate, rather than the performance of each candidate.

References

A Simple Frequency List using R. (2011, April 08). Retrieved October 09, 2016, from http://johnvictoranderson.org/?p=115

Lab 9: Text Data in Tableau. (n.d.). Retrieved October 09, 2016, from http://www.science.smith.edu/~jcrouser/SDS136/labs/lab9/

Peters, G., & Woolley, J. T. (2016, September 26). Presidential Candidates Debates: Presidential Debate at Hofstra University in Hempstead, New York. Retrieved October 09, 2016, from http://www.presidency.ucsb.edu/ws/index.php?pid=118971